Educational Research Bulletin

NOVEMBER 14, 1956
Vol. XXXV, No. 8

A Survey of Users of the Dale-Chall Formula

By JEANNE S. CHALL

THIS survey of the use of the Dale-Chall formula was conducted by the devisers of the formula[1] to satisfy our curiosity. Quite frankly, we wished to know whether the formula had been used, who used it, and what it was used for. Also, we wished to collect evidence on the limitations of the formula. Our own experience with it pointed up certain weaknesses, and we wanted to know about the limitations discovered by other users. We hoped that suggestions from the "ultimate consumers" would help us to improve the formula when we revised it.

The need for such a survey became particularly evident when the educational literature was searched for reports on applications of readability. Compared to the published articles on methods of measuring readability, those on applications were meager.[2] In readability, as in other fields of educational research, the major effort had been made in reporting the discovery of new relationships and techniques. Not enough effort had gone into reporting how these techniques were used, and whether they were helpful. Yet, if educational research is to fulfill its function, its applications should be followed up systematically, and reported in the literature.

The findings of this survey were reported in greater detail in a larger study.[3] They are presented here, although they are dated, because the Dale-Chall formula appears to be used more widely now than it was in 1949. An examination of the latest advertising brochures of large educational publishers show that

[1] The writer collaborated with Edgar Dale in devising the formula which is known by our hyphenated surnames.

[2] For a summary of the fundamental research in readability and applications reported in the literature, *see* Jeanne S. Chall, *Readability: an Appraisal of Research and Application* (Columbus, Ohio: Ohio State University. In press).

[3] Chall, Jeanne S. "Readability: an Appraisal of Research and Application." 1952. An unpublished Doctor's dissertation on file in the library of Ohio State University.

at least four companies specifically mention using the Dale-Chall formula for testing their material. Others refer to the use of readability tests in general. With wider use of the formula there are indications, too, that some users have misunderstood its purpose.

These findings gathered in a survey of the early use of the formula are presented to increase the understanding of readability techniques, and we hope will stimulate the users of formulas to publish their findings.

THE EDUCATIONAL RESEARCH BULLETIN, in January and February, 1948, carried two articles reporting the research which preceded the devising of the Dale-Chall formula for predicting readability and instructions for its use. At that time the staff of the Curriculum Division of the Bureau of Educational Research, Ohio State University, had been using the formula experimentally for about a year. After the articles appeared, copies were sent to 25 persons who were interested in it—most of them were working in research in vocabulary, reading, or readability. By June, 1948, the two articles were reprinted in a pamphlet[4] because the supply of surplus copies in the BULLETIN office had been exhausted.

When orders for the pamphlet began to come in, the mailing department was asked to send a duplicate of each invoice to us. Thus, we hoped to learn who requested the formula. By June, 1949, when this survey was begun, all the pamphlets (1,000) were gone. About 240 invoices or letters requesting the formula were available; these were used for the survey.[5]

The following questions were incorporated in a letter which, with minor variations, was addressed to each of the 240 persons who had ordered the pamphlet:

Who sent for the formula?

Did the person, or persons, who sent for a copy use it?

How was it used?

Were other formulas used? If so, which formulas?

What were the user's reactions to the Dale-Chall formula? What suggestions did he have for improving it? What were his reactions to formulas in general?

[4] The pamphlet received a fair amount of publicity. Announcements and reviews appeared in the *Newsletter* (Ohio State University) and in one or two educational journals. *Editor and Publisher* ran a story on it in July, 1948.

[5] Since many invoices requested several copies (1 invoice covered 100 copies), we felt justified in assuming that most of the 1,000 copies were accounted for by the 240 invoices.

THE largest number of requests came from universities—faculty members, research directors, administrators, and editors of agricultural-extension publications. The next largest number of requests came from the public schools—from superintendents, supervisors of reading, curriculum officers, school principals, directors of research in state departments of education, and the like. The third largest group was designated "individuals." This group included all persons who gave a private address. However, nearly half of those who replied were graduate students and an additional one-fourth were probably teachers. The remaining nine groups are briefly described here:

Book publishers
: Publishers of books and other materials such as pamphlets, tests, and the like, excluding newspapers and magazines

Newspapers and magazines
: Daily and Sunday papers, school newspapers, trade and farm journals, and house organs

Industrial firms
: Large manufacturers, public-relations firms, advertising agencies, employer associations

Armed forces
: Various divisions of the Army, Air Force, and Marine Corps

Religious groups
: Church councils, church missions, theological seminaries, and pastors

State and Federal governmental agencies
: Department of Health, U.S. Department of Agriculture, state civil-service commissions, excluding state departments of education which were included in the category, "the public school system."

Educational, social, and cultural organizations
: Social agencies, unions, youth groups, and the like

University libraries and public libraries

From Table I and these descriptions of affiliations, it is apparent that, although the Dale-Chall formula was published in an educational periodical and at the time of the survey had received few notices outside educational literature, the requests came from fairly varied sources. At least one-fourth of the requests came from people affiliated with newspapers and magazines, the armed forces, religious organizations, state and Federal governmental agencies.

Replies were received to 64 per cent of our letters of in-

quiry—154 replies to 240 letters. Half or more of each group, except the public schools, replied to the letter. The small return by the members of public school systems may have resulted from the fact that many of them were on vacation during the time of the survey. However, since this group made least use of the formula, the lower return may be indicative of a trend. This will be dealt with in detail later. In general, 154 replies, over-all and by groups, seemed to be sufficiently representative of the 240 requests to justify study of them.

TABLE I

THE AFFILIATIONS OF THE 240 REQUESTS FOR THE DALE-CHALL FORMULA

	Number	Per Cent of Total Requests
(1)	(2)	(3)
Universities	65	27
Public school systems	41	17
Individuals	31	13
Book publishers	18	8
Newspapers and magazines	16	7
Industrial firms	14	6
University libraries	13	5
Armed forces	12	5
Religious groups	11	5
Public libraries	7	3
State and Federal governmental agencies	6	2
Educational, social, and cultural organizations	6	2
Total	240	100

WHO used the formula? Ninety-five of the 154 persons who answered our inquiry, or 61 per cent, reported making some use of the Dale-Chall formula.[6] Table II shows the numbers and percentages of persons within each group who reported using the formula. From it we see that within each category, excepting the public schools, at least half of the people who replied reported making some use of the formula. The largest proportion of use was found among the armed forces, where all of the eleven who had replied reported using the formula.

The uses of the formula varied and were divided into the following general classifications: analysis of materials (both published and unpublished), research, teaching, guide for editing and re-writing, general guide for writing, and reference. Table III

[6] This survey was not concerned with the use of the Dale list of 3,000 words apart from the formula. Mr. Dale's list had been available in mimeographed form for at least ten years before it was incorporated in the readability formula.

TABLE II

NUMBER AND PERCENTAGE WITHIN EACH CATEGORY WHO REPORTED USING THE DALE-CHALL FORMULA

	Number of Users	Number Who Replied	Per Cent Users of Those Who Replied*
(1)	(2)	(3)	(4)
Universities	26	47	55
Armed forces	11	11	100
Individuals	10	18	56
Industrial firms	9	11	82
Book publishers	9	15	60
Newspapers and magazines	7	11	64
University libraries	6	7	86
Religious groups	5	8	63
Public school systems	4	13	30
Public libraries	4	6	67
State and Federal government agencies	2	3	67
Educational, social, and cultural organizations	2	4	50
Total	95	154	61

*Within each category.

TABLE III

UTILIZATION OF THE DALE-CHALL FORMULA BY THE 95 PERSONS WHO REPORTED SOME USE OF IT

Types of Use	Number of People*	Per Cent of Total Users*
(1)	(2)	(3)
Analysis of manuscripts	48	51
Research	23	24
Teaching	19	20
General guide for writing	19	20
Editing and re-writing	12	13
Reference	7	7

*Both columns total to more than 95 and 100 per cent because more than one use was mentioned by many people.

indicates the number of people, within all groups combined, who reported using the formula for the indicated purposes (*see* Column 1).

The table shows that the formula was used most often for analyzing materials. All groups used it for this purpose. Research was the second most frequently mentioned: these users came mainly from three groups—universities, individuals, and armed forces. An equal number reported using it as a guide for writing and in teaching. Those who used it in teaching came mainly from the universities. However, the formula was also "taught" by religious groups, industrial firms, and state and federal agencies. Some people in almost all groups used the formula as a

guide for writing. Book publishers, and newspapers and magazines, together, made up about one half of those who mentioned this use. Book publishers also used the formula in editing and re-writing. The least frequent use of it, that is, for general reference, was mentioned only by the universities and public libraries.

OTHER formulas were frequently used in conjunction with the Dale-Chall formula. Fifty of our 154 respondents specifically mentioned using at least one other formula. The other formula most frequently mentioned was the Flesch, which was named by at least one person in nearly every group. A total of 40 respondents mentioned it. The next most frequently mentioned formula was the Lorge—12 people among book publishers, public school systems, and universities told of using it. Five or fewer people mentioned the Winnetka formula, the Yoakam, the Gray-Leary, the Lewerenz, the Edgerton formula (a modified Winnetka technique), the Gunning formula, individual simplifications of existing formulas, and various word lists such as those of Rinsland and Thorndike.

The survey was not meant to be a popularity contest. Indeed, the opinions of our respondents could not be considered representative of the general run of formula users at the time. However, the following information comparing the different formulas is of general interest in light of statements often appearing in the literature that are based on the opinions of an individual author's experience with one or two formulas.

How did the Dale-Chall compare with other formulas? Very few of the respondents had used several formulas sufficiently to be able to give us reliable information in answer to this question. Still fewer thought that they wanted to commit themselves. However, 22 respondents compared the Dale-Chall with other formulas. Sixteen stated their preference on the basis of ease of application: ten preferred the Dale-Chall; five, the Flesch; and one, the Lorge.

The findings are particularly interesting in light of others which have been reported. Yoakam and his students, for example, have consistently found the Yoakam formula the easiest one to apply.[7] Lostutter[8] has reported that he found the Flesch formula

[7] Yoakam, Gerald A. "Reading Difficulty of School Textbooks," *Elementary English Review*, XXII (December, 1945), pp. 304-309.

[8] Lostutter, Melvin. "Some Critical Factors of Newspaper Readability," *Journalism Quarterly*, XXIV (December, 1947), p. 310.

easier to apply than the Lorge. Gunning, on the other hand, writes that the Dale-Chall is the easiest to apply.[9] We think that the only conclusion that can be drawn from these varying reports (including our findings) is that familiarity with a formula is probably the most important factor in ease of application.

Six people mentioned preference on some other grounds than ease of application. One publisher preferred the Dale-Chall for Grades IV, V, and VI. One school preferred the Dale-Chall to the Lorge for use in high school and college. One public library preferred the Dale-Chall to the Flesch for analyzing juvenile books. One publisher preferred the Lorge to the Dale-Chall because it gave more exact grade-level designations. A writer of children's books preferred the Flesch to the Dale-Chall for Grades IX and X. A university professor in England preferred the Flesch because its lack of a word list made it more transferable to use in England.

IN GENERAL, the users reported satisfaction with the Dale-Chall formula. The great majority found it useful for the purposes to which they put it. Only 17 people, many among the "satisfied users," offered suggestions for improving it.

The most frequently mentioned weakness of the formula was the word list. Six people questioned its validity for estimating difficulty of materials in a special field of knowledge for an audience familiar with the field. They suggested adding specialized word lists to make the formula more valid for these special conditions. In fact, several people had already begun their own specialized lists—one of farm[10] words and the other of army terms.

Three book publishers thought that the grade-level designations of the Dale-Chall formula were too broad. They felt our corrected grade-levels of fifth-sixth grade, seventh-eighth grade, and the like were not exact enough for their purpose.[11] One of the book publishers also suggested that the formula would be more valuable if it could be applied to books below the fourth grade. This cannot now be done since the Dale word list of 3,000

[9] Gunning, Robert. *The Technique of Clear Writing.* New York: McGraw-Hill Company, 1952, p. 33.

[10] This study has been completed recently and is available in a pamphlet by Lloyd Bostian and Bryant Kearl, *Frequency of Economic Terms and Concepts in Farm Magazines,* published by the University of Wisconsin Press (Bulletin No. 25, Department of Agricultural Journalism, College of Agriculture).

[11] We used the rough grading to make realistic the lack of precision that the formula actually has. However, in spite of the caution about the approximate nature of the measure, many would like it to be more precise even if it perhaps violates validity.

words contains words known by fourth-graders. The constant in the formula is 3.6365, thus putting all material on at least fourth-grade level.[12]

Other points mentioned by only one or two individuals, but of interest in application of the Dale-Chall formula, follow:

Better rules on the repetition of words are needed. As the rules now stand, unfamiliar words are counted unfamiliar as often as they appear. Since some materials repeat the same words (mostly technical words) many times, an unusually high grade-level may result.

The word list should be brought up to date. Children now know the words: *atom*, *television*, and the like, which are not on the list.

Better rules regarding names are needed. The respondent mentioning this point (a publisher) thought that by counting biographical and geographical words familiar, we were actually underestimating the difficulty of social-studies materials.

Some evidence is desirable on appraising the difficulty of poetry.

Rules on hyphenated and compound words should be improved.

In these summary items an attempt is made to quantify the information contained in the 154 replies to the inquiry. To regain some of the individual flavor, and to get a better understanding of the "readability situation" among the different groups in 1948-49, a more qualitative analysis of the responses follows.

THE largest group of users of the Dale-Chall formula, people associated with universities, used it mainly for instruction and research. They also used it for estimating the difficulty of books and other materials. Since a very small number used the formula for evaluating their own writing, none reported using it for editing and re-writing. However, several used the general principles as a guide for writing. Universities used other formulas beside the Dale-Chall. Those mentioned were the Flesch and the Lorge.

The following summary will give the reader some idea of the various uses to which the Dale-Chall formula was put:

Instruction

It was discussed or taught in regular classes in psychology, journalism, communication, educational research, or in writing workshops.

[12] George Spache of the University of Florida, Gainesville, has since published a formula for the primary grades. *See* "A New Readability Formula for Primary-Grade Reading Materials," *Elementary School Journal*, LIII (March, 1955), pp. 410-13.

Research

The Instructional Film Research Center, Pennsylvania State College, used the formula to equate verbal difficulty in a film experiment to find the optimum number of facts to be presented in instructional films.

The Motion Picture Research Project, Yale University, used it to control the difficulty of sound tracks on various films. The same film was used with both an easy and a difficult sound track (as pre-tested by the formulas). Both versions of the film were shown to groups of people to determine the effect of the difficulty of the sound track on the general effectiveness of the film.

At the University of Minnesota it was used to study the influence of readability on readership of newspapers.

Analysis of Manuscripts

Some of the materials analyzed were:

Radio scripts for health and other programs for children at the University of Texas.

College textbooks at San Diego State Teachers College.

Publications of public agencies, newspapers, and other media at the Division of Research Interpretation, Institute for Research in Social Studies, University of North Carolina.

Books for collateral reading, textbooks considered for adoption, manuscripts before publication at Stephens College.

Agricultural-college press releases at the Department of Agricultural Journalism, University of Wisconsin.

Textbooks and periodicals in health and physical education at the University of Illinois.

Children's fiction, children's textbooks, comics, children's magazines, government publications for mass reading at the University of Birmingham (England).

THE most outstanding feature about the schools, as a group, when compared with the other groups, is that they had the lowest proportion of replies and also reported the least use of the formula. Of the thirteen who answered our inquiry, only four said that they had used the Dale-Chall formula.

Since the majority were non-users, let us examine this group first. Of the nine who did not use the Dale-Chall, seven said "not yet" instead of the usual "no" in answer to our question, "Did you use the Dale-Chall formula?" Why not yet? Their answers all implied *no time*.

The various school people—chiefs of textbook and publications committees, curriculum assistants, and other school personnel had intended to use the formula for checking on the difficulty of textbooks for state adoption, for placement on booklists, or for classroom use. However, in each instance they said, in essence: We want to use it, we know it is important; our teachers know it is important, but we lack the time, and no clerical help has yet been provided for it.

A study reported in 1951 by the Committee on Diagnostic Tests, entitled *Reading Instruction in the Public Schools*, also found a very small proportion of teachers using readability formulas to judge the difficulty level of textbooks. Most of the teachers in this survey reported using publishers' grade placements of books or making their own judgments of the difficulty of the vocabulary. Although the authors of readability formulas have presented their techniques for the use of teachers, the evidence seems to indicate that formulas will no doubt remain, for some time to come, a practical tool for authors and publishers, a research tool for university professors, and a "neat subject" for Masters' and Doctors' theses.

Those who used the formula analyzed textbooks for use in a reading clinic, checked the difficulty of manuscripts of books written for specified school grades, and used it as part of a general in-service teacher-education program. The most comprehensive use of the Dale-Chall formula was reported by the Oakland public schools. There the Dale-Chall formula was used as one of several devices for making secondary-school administrators and teachers aware of the reading problems brought about by population changes in the community during the war.

Various groups were taught the use of the readability formula. These groups included principals, librarians, and curriculum assistants of all their secondary schools, teachers of science, industrial arts, and English. They found it took about 50 minutes to explain the formula to a group and have the individuals work the reading levels of two samples from books in their own fields.

In teaching the formula, they pointed out: What the formula will not do; what it will do. They discussed factors which make for reading difficulty, and emphasized the fact that the true measure of a reading device is in its actual use in the classroom.

Using the formula as a tool in adjusting methods and materials to a changing school population had many beneficial

effects. Some of the specific benefits mentioned in their lengthy report were: more money was spent on needed simple reading materials, teachers ordered small assortments of books on different reading levels to meet the range within a class, they listed several approved textbooks on different reading levels in a single field, public-relations bulletins addressed to parents were simplified so that they could be read easily by most people. They also compiled lists of materials (textbooks, supplementary books, pamphlets, charts) successfully used in the different grades in biology, chemistry, physics, physiology, and other fields, by teachers' judgments of interest and reading level, and by the Dale-Chall reading level. The Oakland public schools' solution of the clerical and financial problems of readability seems a realistic one. By involving teachers and administrators, and acquainting them with readability techniques, the tremendous job of appraising books was not only divided among them, but they learned the limitations of the formula.

In summary, few of the public schools that requested copies of the formula used it. Although most of the respondents sensed the need for using readability measurement, lack of time, clerical help, and the necessary funds kept them from carrying out plans for readability evaluations. The Oakland public schools, however, solved this problem by using the Dale-Chall formula as part of a general attack on their reading problems. Groups of secondary-school teachers and administrators were taught to use the formula, and co-operated in evaluating books widely used in the school system.

Most of the publishers reported making some use of the Dale-Chall formula. All of the users made analyses of either published works or manuscripts. The materials analyzed were: textbooks in reading, arithmetic, and social studies; vocational pamphlets; psychological tests; and selections from a children's encyclopedia. The books were intended for children in Grades IV through VIII. The formula was also used for editing and rewriting. Four specifically mentioned making some adjustments in manuscripts to bring them in accord with formula results.

Three publishers reported using the formula as a general guide for writing. One used it indirectly in setting up standards on sentence length. Another used the Dale list of 3,000 words as a guide for choice of new words in primary reading books. An-

other used the general principles in writing pamphlets containing vocational information and devising psychological tests.

Compared to all other groups, the publishers had the most experience with readability formulas. All but two of the nine users of the Dale-Chall formula specifically mentioned using at least one other formula or word list for checking reading difficulty. In fact, most of the publishers reported using at least three other formulas or word lists. The formulas in the order from most frequent to least frequent mention were: Lorge, Flesch, Winnetka, Yoakam, Gray-Leary, Lewerenz, and Edgerton (a modified Winnetka technique). Only three offered comparisons among the formulas. All three found that the various formulas gave results in sufficient agreement with one another.

Several publishers made suggestions for improving the formula for their purposes. Three suggested narrower corrected grade-levels, and one suggested that names of persons and places be considered "unfamiliar."

Two publishers expressed their opinions about readability formulas in general. Both thought that much more experience in the use of them was needed. However, the formulas were used because school people wished to know the levels of the books they bought, and according to one editor of a large textbook publishing house, at least one state required publishers who submitted textbooks for adoption to supply information concerning the readability of the textbook in question. (A formula was to be used although no particular formula was prescribed for this purpose.) This editor also suggested that the deviser's explanation of a formula should give pertinent information to caution the user against applying the formula to materials for which it was not intended. More specifically, he suggested the addition of certain common words in special subject areas (arithmetic, science, and so forth) to make the formulas applicable to these special materials as well as to general reading matter.

Among the representatives of newspapers and magazines who replied to our letter, we found general acceptance of the principles growing out of readability measurement, but limited use in the testing of materials. For example, only seven of the 16 persons who answered our inquiry, reported making some use of the Dale-Chall formula. Two of them, however, used it to test the readability of their newspapers or magazines. Four said they

read the pamphlet for its general principles and used them as a guide for writing. One magazine editor reported using the Dale-Chall formula for research in a split-run experiment relating readability to reader interest. Four also reported using the Flesch formula, and two had regular analyses of their magazines done by Robert Gunning. One editor,[13] a pioneer in the application of readability measurement, wrote that he had been using readability measurement since 1946, making the greatest use of the Flesch formula. As a matter of editorial routine, they worked readability scores on samples of articles, passed them around the office, and got the staff writer in the habit of writing with these standards in mind. He also noted that the use of split-run, reader-interest surveys had a considerable effect on the staff. After a split-run survey indicated that a more readable score would push up the reader-interest score, "there was more inclination to watch these factors in the writing of new copy."

Our survey indicated that readability formulas were not so widely used among newspaper and magazine publishers as one would expect. While the general principles were being applied, according to our respondents, very few undertook extensive analyses of their publications with formulas. When such analyses were undertaken, the Flesch formula was commonly used.

Our general interpretation was confirmed by a letter from the American Press Institute, Columbia University. They had discussed the Dale-Chall formula at the institute, but had not yet used it. They did, however, occasionally use the Flesch formula, although few editors used readability formulas persistently: "A large number of them now accept ideas growing out of use of the formulas but are inclined to follow rule-of-thumb methods instead of going to the trouble of counting words and doing other things that use of formulas requires."[14]

THE most frequent use of the formula by industry was in estimating the difficulty of management newsletters, radio copy, advertising copy, management editorials, and questionnaires to employees. Eight firms reported making such analyses of their

[13] Donald R. Murphy, editor of *Wallace's Farmer and Iowa Homestead*. *See* "How Plain Talk Increases Readership 45% to 66%," *Printers Ink*, 220 (September 19, 1947), pp. 35-37.

[14] For descriptions of readability campaigns sponsored by the United Press and the Associated Press, *see* Robert Gunning, *The Technique of Clear Writing* (New York: McGraw-Hill Company, 1952), and Rudolph Flesch, *The AP Writing Handbook* (New York: Associated Press, 1951).

printed materials. The results indicated, as one respondent wrote, "What should have been suspected—that the information was over the heads of most of the people it was intended to reach."

The public-relations director of a large steel firm reported using the Dale-Chall formula to analyze materials for teaching-aid purposes—periodical material to be sent to science teachers and students in high schools. They attempted to gear the material to seventh-grade readability for easy understanding by junior-high school students as well. They also hoped to use some of the material for industrial-relations and vocational-training courses and in the training of steel salesmen. Several firms also used the formula for editing and re-writing. Two firms reported using the formula in their general approach to writing new materials.

Five firms mentioned using some other formula, all of them the Flesch. These five were also users of the Dale-Chall. No other formulas were mentioned.

Three individuals made some suggestions for improving the formula. All specifically mentioned the value of having additional specialized word lists. One suggested that better provision be made for the repetition of common technical words. Both the Flesch and the Dale-Chall formulas consider the words difficult every time they are used. Since a short technical booklet may repeat an important word as often as 200 times in 48 pages, it was suggested that some adjustments therefore be made.

In general, industrial users of the formula appeared to be aware of the difficulty of their printed materials. They were using readability formulas to check the difficulty of their sales literature, their advertising copy, employee communication, questionnaires, and teaching-aid materials that are distributed among schools. They were also re-writing some of the materials that were too difficult and were also using the generalizations from readability for writing new materials.

THE armed forces were the most extensive users of the Dale-Chall formula. The chief use of the formula was for research. Other uses were the analysis of materials for selection in reading programs, as a guide in editing and re-writing, and in original writing. The kinds of research studies in which the Dale-Chall formula was used were:

To compare the readability of selected training literature produced by the U.S. Air Force during the Second World War with the read-

ability of representative training literature produced by the War Department shortly before the war.
To pre-test directions for a classification test battery.
To determine the reading difficulty of army field technical manuals.[15]

The most frequent use of the formula was for selecting materials for the Army reading programs. At the reading laboratory, Tyndall Air Force Base, Florida, the formula was used to grade books of fiction and non-fiction. A total of 135 books was analyzed. Ninety-four of them ranging in difficulty from fifth- through eighth-grade level were selected for training purposes. They were used with the reading rate controller to speed up reading. The titles of 41 books, ranging in difficulty from ninth through fifteenth grade, were listed in a pamphlet, *A Guide for Professional Reading by Air Force Officers.*

At the National War College the formula was used for book evaluation and for test construction, in connection with the reading-improvement program. Books at certain levels of difficulty were used for improving rate of reading. In test construction, the formula was used to equate the difficulty of various selections. They had to be the same readability level to determine either student progress or the effectiveness of the course in so far as reading rate was concerned.

At the Air University, Maxwell Air Base, the Dale-Chall formula was used to select materials for their reading-improvement program. Numerous instructional materials—general and technical—were graded with the formula.

Several people in the armed forces used the formula for editing, re-writing, and original writing. Four of the respondents from the armed forces mentioned using the Flesch formula. No other formula was mentioned.

Several of the respondents from the armed forces suggested some improvements for the formula. First in importance, was the need for a specialized word list. Two persons specifically mentioned this and one actually constructed one for his purpose.

Most of the people affiliated with religious groups who answered our inquiry used the Dale-Chall formula for checking the difficulty of various books, pamphlets, sermons,

[15] This research was carried out by Joseph E. Melanson, 1st Lieutenant, U.S.A. It was submitted as a Master's thesis at the University of Wisconsin entitled "The Dale-Chall Readability Formula Applied to Army Field, and Technical Manuals." Bryant Kearl directed the study, in which 139 army terms were added to the Dale list of 3,000.

and the like, as a guide for re-writing and in writing new material. It was also taught at workshops.

It was used in a project on curriculum development at the Protestant Episcopal Church, Greenwich, Connecticut. The formula was used to test books written for children.

The staff of the Sunday schools of the Lutheran Church, Missouri Synod, used the formula as a quick check of the readability of their various types of pamphlets and lesson materials.

Sermons and other religious writings were analyzed and simplified. The formula was also used as a guide in re-writing several difficult theological papers for an adult study course. One rector used the formula to convince a psychiatrist that he needed to simplify his writing on mental hygiene for the use of the clergy.

The formula was also used in writers' workshops at the Chicago Theological Seminary. Writers analyzed and evaluated the communication qualities of whatever program they wrote. The director noted that they

> . . . got some fairly startling learning on the part of individuals out of such analyses. I know of no quicker way by which ministers can become aware of the stilted and involved language which they use. I might add that we found it quite important psychologically to let the writer do his own computations and evaluations.

[Vol. XXXV, No. 8]

* * *

THE "position section" of the mailing list of the EDUCATIONAL RESEARCH BULLETIN is being checked. To explain the "position section" I need to look back to the beginning of the BULLETIN in 1922. At that time the editor decided, as a public service, to send the magazine to the position "superintendent of schools" in each city, town, and village with a population of twenty-five hundred or more in the country. He did this hoping that the BULLETIN might aid in developing a nation-wide interest in the new activity called "educational research." The cost of printing has risen to such a figure that the Director of the Bureau of Educational Research has asked that persons who wish to continue receiving the BULLETIN verify their addresses in the mailing list. We gave the date November 15 as the final date. Since cards are still coming in, we shall make no change until after the December issue is mailed. If you still want to receive the BULLETIN, be sure to send in your verified address.

An Evaluation of Two Methods of Grouping

By DARRELL HOLMES
AND
LOIS HARVEY

It is axiomatic that children differ. All teachers seem to agree that provision must be made for these differences. The methods utilized to organize a class to provide for individual differences vary with the subject, the children, the teacher's preference, and numerous other factors. In the teaching of arithmetic a number of approaches have been suggested; included among these are the plans of permanent and flexible grouping. Data concerned with the relative effectiveness of these two methods have been gathered; and the presentation and evaluation of these data constitute the substance of this paper.

For the purpose of this study, *permanent grouping* is defined as that plan or method by which the teacher divides his class into two or more groups. This is usually done on the basis of information gathered at the beginning of the year. The groups are maintained throughout the year, with very little shifting of children from group to group. These groups presumably are taught almost as separate classes. *Flexible grouping* is that plan or method in which a teacher teaches the class as a whole whenever new materials are introduced, and then divides the class into groups on the basis of accomplishment in this particular topic. Those attaining high levels are placed in one group and others not doing so well are placed in another group or groups. In the latter groups, the materials are reviewed and otherwise retaught; in the former group, provision is made for practice and enrichment.

The present study consisted of two parts. Initially, the writers were concerned only with the question of subject-matter attainment. Two third-grade and two fourth-grade classes were involved in this phase of the study. In so far as was possible, the classes selected were similar with respect to intelligence quotients, socioeconomic background, and the ability of their instructors. Aside from using different methods of grouping, the approach in teaching was that of emphasizing meaning and practice. Arithmetic tests were administered to the four classes on October 20 and again on March 4.

The results obtained suggested certain conclusions which are presented in subsequent sections of this paper; however, as the

subject-matter outcomes were being evaluated, questions concerning the children's attitudes toward arithmetic and social development were increasingly important. As a consequence, the second part of the study was conducted. In this phase, two sixth-grade classes were selected according to the same standards used with the third- and fourth-grade classes. The sixth-grade classes were tested in October of the following year and again in May (the time elapsed was approximately eight months). Not only was a standard arithmetic test used in these classes, but also an aptitude test and a sociometric scale were administered. Thus, it was possible to gather data regarding subject-matter changes by using 168 subjects, and data regarding the attitudes and social adjustment by using 63 subjects. It should be noted that initial plans called for the involvement of more than the six classes utilized in the study, but circumstances beyond the control of the investigators made this impossible.

As has already been indicated, an effort was made to select classes which were similar with respect to student ability, socioeconomic background, and the teaching effectiveness of the teacher. This was essentially a matter of judgment in selecting already existing classes, because it was impossible to select individuals and to make up special experimental classes. Certain after-the-fact observations about the nature of the classes demonstrate that the classes were comparable.

The third-grade classes were in the same school building in a federal-housing district in San Diego, California. Most of the parents either worked in one of the local airplane factories or were connected with the Navy as civilian workers or enlisted personnel. Neither class had been together as a unit during the previous year; each was made up of children from different classes within the school, and from other schools, districts, and states. Children in each class were from several races. The fact that about 25 per cent of the children in each class had moved out of the school area by the beginning of March in the year of the experiment, with additional children moving in, indicates that comparable factors affecting mobility were operative in each class. Moreover, when intelligence-test scores became available, it was noted that only a one-point difference existed between the two classes. The distributions of intelligence quotients were tested and found to be comparable. All of this evidence lends support to the assumption that the third-grade classes were essentially

matched groups. Comparable observations were true for the fourth-grade classes.

The two sixth-grade classes were drawn from different schools, but these schools served similar areas. When a final check was made of the intelligence quotients, a mean difference of 6 points was found in favor of the class utilizing the permanent groups. This could have worked to the advantage of the permanent groups; actually, as can be noted in the section on findings, it did not.

The third-grade teacher using permanent grouping and the fourth-grade teacher using flexible grouping were convinced as to the value of the methods of their preference.[1] The other four teachers had no strong opinions about the relative merits of either method of grouping. Teachers were rated good by the supervisors and administrative personnel; they varied considerably in teaching experience. Four of the teachers were women; men were found in the fourth-grade and sixth-grade flexible classes.

The standardized achievement tests used in testing arithmetic were those which were normally a part of the school's program. In the third grade, two forms of the Stanford Primary (old form) were used; in the fourth grade, the Intermediate Battery of the Metropolitan Achievement Tests. In the sixth grade, the arithmetic section of the Intermediate Battery of the Stanford Achievement Tests (new form) was used.

To assay changes in attitude toward arithmetic, a scale was developed with the co-operation of William Stegeman, director of research, San Diego city schools. This scale was entitled, "What Do You Think?" and consisted of ten items from which the following is taken:

1. If your class has been practicing for a play and there isn't enough time for all the lessons, which would you like to have left out?
1——Reading 2——Arithmetic 3——Music 4——Spelling

This test was constructed to be scored by counting the number of times arithmetic was selected in a favorable context and subtracting from it the number of times it was selected in an unfavorable context.

A sociogram was also constructed with the assistance of Mr. Stegeman. It involved three parts: The child was asked to

[1] It is not known how these differences among the teachers affected the results of the experiment. It was the judgment of the present writers in advance of conducting the study that errors introduced into the experiment would be of a random nature, and therefore, should not affect the results in a consistent manner.

name in order of preference: (1) the three children with whom he would like to work on a new problem in arithmetic; (2) the three children with whom he would like to serve on a committee; and (3) the child to whom he would like to give one of his two pieces of candy. The actual methods used in evaluating the results of the sociogram are discussed later in this report.

TABLE I

SUBJECT-MATTER RESULTS

GRADE AND GROUPING	NO. OF PUPILS	MEAN IQ	ARITHMETIC			
			Pre-test	Post-test	Difference	t
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sixth Grade—						
Flexible	32	101	5.65	6.55	.90	10.00
Permanent	31	107	6.00	6.84	.84	5.42
Difference between differences					.06	1.06
Fourth Grade—						
Flexible	32	100	4.13	4.63	.50	9.26
Permanent	31	101	4.00	4.23	.23	5.38
Difference between differences					.27	4.42
Third Grade—						
Flexible	21	99	2.79	3.32	.53	6.27
Permanent	21	99	2.79	3.51	.61	7.31
Difference between differences					−.08	−.68

IN TABLE I data are presented which reveal the relative effectiveness of the two methods of grouping with respect to subject-matter outcomes. The type of grouping, mean intelligence quotients, number of students involved, and pre- and post-test scores are presented in this table. In the third grade, gains of .53 (months) and .61 (months) are found for the flexible and permanent groups, respectively. These represent highly reliable gains because the respective t's are well beyond the .01 level of significance. Note, however, that the difference between the gains is .08; this is not statistically reliable. At the sixth-grade level, the difference of .06 between the mean gains could also have occurred by chance. However, the difference of .27 in favor of the flexible class in Grade IV is unlikely to have occurred by chance; hence, the problem remains of evaluating its meaning. Actually, by no stretch of the imagination could it be possible to attribute the obtained difference to the method of grouping. In

TABLE II

GAINS IN ARITHMETIC ANALYZED ACCORDING TO INITIAL KNOWLEDGE OF ARITHMETIC

GRADE AND GROUPING	PERMANENT		FLEXIBLE		DIFFERENCE OF DIFFERENCES
	Number of Pupils	Gain	Number of Pupils	Gain	
(1)	(2)	(3)	(4)	(5)	(6)
Sixth Grade—					
High*	15	.58	15	.81	−.13
Low†	16	1.10	18	.92	.18
Fourth Grade—					
High*	23	.20	21	.45	−.25
Low†	8	.29	11	.60	−.21
Third Grade—					
High*	9	.62	8	.37	.25
Low†	12	.45	13	.63	−.18

* At or above grade-level in arithmetic on the pre-test.
† Below grade-level in arithmetic on the pre-test.

the first place, though every effort was made in advance to select teachers who were comparable with respect to the important variables in teaching, it was noted by the supervisor that considerable difference in enthusiasm existed between the two fourth-grade teachers. The teacher using the flexible method was quite enthusiastic about flexible grouping. In additon other considerations enter into the picture. The teacher of the permanent group stated quite definitely that she was going to have to change the method of teaching because of the nature of the class. In the writers' judgment, then, the obtained difference between the two classes is due to factors other than the kind of grouping used.

In the light of the evidence presented as evaluated in terms of the experimental situation, it is concluded that, when considering over-all results, the method of grouping used makes very little difference with regard to promoting subject-matter outcomes. Each method appears to be equally good.

CONSIDER next the relationship between prior knowledge of arithmetic and gains effected. It has been asserted that the method of permanent grouping tends to heighten differences in knowledge of arithmetic. Is this in fact borne out by the data? For purposes of analysis each of the six classes was divided into two groups: one group at and above grade-level in the pre-test

groups, and the other group below this point.[2] Average gains in arithmetic were determined in each class for each of these groups; these are presented in Table II. Thus, as can be noted in this table, the average gain for the sixth-grade permanent group, high in prior knowledge of arithmetic, is .58, and in the flexible group it is .81. Note that the gain for the sixth-grade permanent group, low in prior knowledge of arithmetic, is 1.10. This suggests that the low group was closing the gap between it and the high group. This is shown by the answers to the following question:

Are increasing differences found between high and low groups?

	Permanent	Flexible
Yes	1	0
No	2	3

In five of the classes increasing differences between high and low groups were not found. This suggests that the enrichment work, however it was being accomplished, was not primarily increasing the difference in knowledge of arithmetic as measured by the test used. It does not appear to matter how the class is organized. Differences in gains of arithmetic, when considered against the background of the student's prior knowledge, apparently depend upon factors other than the method of grouping involved.

In Column 6 of Table II differences are presented for comparable groups in the two plans of grouping. For example, note that the difference between the high, sixth-grade groups is —.13, and that the minus sign indicates that the difference is in favor of the flexible group. This in effect enables comparisons to be made between the two methods of grouping to determine whether or not one method promotes more learning in a specific subgroup than the other.

Ignoring the fourth-grade classes in which the teacher variable is quite likely to have caused the differences, it is noted that the differences in gains do not show a consistent pattern in favor of either group. It is concluded, therefore, that the method of grouping did not significantly affect the learning of arithmetic when groups of comparable aptitude were compared.

ANOTHER question which comes to mind deals with the interrelationships of learning ability as measured by intelligence tests and the method of grouping in promoting subject-matter

[2] The interpretation would have been much more straight-forward and meaningful had it been possible to use an advanced statistical design. This was not possible because of the necessity of fitting the experiment to the situation rather than *vice-versa.*

attainment. It is recognized that intelligence-test scores are imperfectly related to accomplishment in arithmetic. On the other hand, if intelligence quotients can be assumed to be indexes of learning ability, it would be meaningful to analyze differential gains among various groups. Data with regard to this point are presented in Table III. Note, for example, that the gain in knowledge of arithmetic in the sixth grade, permanent class high I.Q. group is .93 and that the gain in the flexible class is .97. The final column in the table gives the differences between the two groups; in the case of the example it is a —.04.

TABLE III

GAINS IN ARITHMETIC ANALYZED ACCORDING TO INTELLIGENCE QUOTIENTS

GRADE AND I.Q. GROUP	PERMANENT		FLEXIBLE		DIFFERENCE
	Number of Pupils	Gain	Number of Pupils	Gain	
(1)	(2)	(3)	(4)	(5)	(6)
Sixth Grade—					
High*	22	.93	17	.97	—.04
Low	9	.62	15	.83	—.21
Fourth Grade—					
High	16	.18	12	.52	—.20
Low	15	.27	20	.49	—.21
Third Grade—					
High	9	.91	7	.49	.42
Low	11	.43	14	.56	—.13

* Intelligence quotients were above 101.

Does the method of grouping seem to cause a disproportionate gain between the several intelligence-quotient groups? This question is analyzed by looking at the various pairs and determining in each class which group gained the most. For example, in the third-grade, permanent class, the high intelligence-quotient group gained .91 and the low group .43. In effect the gap between these two groups in knowledge of arithmetic widened during the course of the class. The total count is shown by the answers to the following question:

Do students in the high intelligence-quotient groups tend to gain more than the students in the low groups?

	Permanent	Flexible
Yes	2	2
No	1	1

Note that in four out of the six classes the high intelligence-quotient groups profited more than the low groups, and that no difference appears consistently in either of the two methods.

This suggests, therefore, that a widening of the gap between high and low groups tends to occur with respect to knowledge of arithmetic. However, if this is true, it is quite likely a function of whatever is measured by the intelligence test, and not of the method of grouping.

Disregarding the fourth-grade results, again because of the teacher variable being the major factor, it can be noted that three out of the four differences (*see* Column 6, Table III) between the two methods of grouping are in favor of the flexible method of grouping. The differences are not large enough or consistent enough to warrant a specific conclusion. A special experiment should be designed to test the hypothesis that flexible grouping promotes more learning in the high I.Q. groups than does the permanent method.

THE sociometric instrument described in a previous section was administered at the beginning and again at the end of the experimental sessions. This instrument asked for selections of leaders, co-workers, and friends. It provided evidence for the analysis of the effect of the method of grouping on social structure and individual social development. The data were analyzed in two ways: first, sociograms were developed; and second, choices were quantified and scores were obtained for each child in each of the selection areas. For example, a child whose name was selected several times as being a leader received a higher score than a child receiving no selections.

The sociograms became quite complex and are not presented in this report. They revealed changes in group structure, but those changes which were identified could not be attributed to the method of grouping. Changes of a comparable order occurred in both classes. Specifically, changes in selection patterns, leadership patterns, and isolate patterns were essentially the same in each class. For example, four individuals in each class received no first-place selections in either the pre- or the post-test situations. Moreover, the total number of isolates identified in each class was exactly the same. One would presume that the method of flexible grouping would encourage more interaction among individuals and a consequent differential effect in social and individual

development. The data in this section do not support this expectation.

In this experiment, no noticeable difference in social development was ascertained in students of the two classes. The method of grouping did not appear to be a crucial factor in the selection (or nonselection) of friends, leaders, or co-workers.

Several other interesting possibilities were also analyzed using the quantitative scores. For example, it is possible that a child who "catches fire" in arithmetic might attract other children as co-workers in arithmetic. An analysis of the data did not support this expectation. In addition, no changes in childrens' sociometric scores could be ascribed to initial knowledge of arithmetic or to intelligence-test scores. The data involved in these analyses are tedious and not worth presenting in tabular form.

The attitude questionnaire described in a previous section of this report was administered to the sixth-grade classes in October and May. Since the gains on five of the ten items were in favor of flexible grouping, those on four were in favor of permanent grouping, and one item was a tie, it seemed obvious that no significant difference exists between the two methods with respect to engendering a favorable attitude within students toward arithmetic.

THE general problem investigated was the differential effects, if any, of two methods of grouping children in arithmetic. Standard arithmetic tests were administered to six classes, two each in the third-, fourth-, and sixth-grade levels. Moreover, arithmetic attitudes and sociometric scales were administered to the sixth-grade class. All tests and scales were administered at the beginning and the end of each part of the study. Data so obtained were analyzed and the following conclusions with respect to the experimental situation were drawn:

The method of grouping used in organizing these arithmetic classes made very little difference with regard to promoting subject-matter outcomes.

Regardless of the methods used, grouping did not increase already existing differences in knowledge of arithmetic. Specifically, students who were at or above grade-level in arithmetic at the beginning of the semester did not necessarily learn more arithmetic than did students who were below grade-level in arithmetic at the beginning of the semester.

The method of grouping did not significantly affect the learning of arithmetic when groups of comparable aptitude were compared. Specifically, neither the flexible nor the permanent method of grouping had the advantage in promoting subject-matter learning in a specific subgroup, as, for example, promoting learning in the group already knowing much arithmetic.

The data suggested that a widening of the gap between high and low intelligence-quotient groups tended to occur with respect to knowledge of arithmetic. This was irrespective of the method of grouping.

The method of grouping appeared to have no effect on the social structure of the class.

The method of grouping appeared to have no differential effect on the attitudes of the children toward arithmetic, though in both classes the attitude toward arithmetic was less favorable at the end of the year than at the beginning.

In general, the data of this study support the conclusion that the method used in grouping arithmetic classes in order to meet individual differences is not particularly crucial. Educational outcomes with respect to subject-matter learned, social acceptance and development, and attitudes toward arithmetic are apparently not related in a systematic way to grouping procedures. It would appear in the judgment of the writers that factors identified in other studies are of primary importance, not the method of grouping. These factors include the personal and professional qualities of the teacher and the teacher's emphasis on meaning and functional practice in his teaching.

Perhaps the findings would be different had the study been conducted in a setting where grouping was done by classes; that is, where all bright students were placed in one class and so on. Moreover, this study has been concerned only with the differential effects of permanent *versus* flexible grouping. No effort has been made to compare these two methods of grouping with other methods, or, in fact, with no method of grouping.

Pending further study of this problem, it is the writers' judgment that each teacher should analyze his own class and then make provision for individual differences in the ways that, in his professional judgment, will best meet the needs of that particular situation. To argue that one particular method is better than the other is unproductive.

READINGS

American Library Association. *Subject and Title Index to Short Stories for Children.* Compiled by a subcommittee of the A.L.A. Editorial Committee, Julia F. Carter, chairman. Chicago: American Library Association, 1955. vi + 333 pp.

This is an index to 372 collections of short stories for children and is designed for the use of public and school librarians and for teachers who work with children from the third grade through junior high school. It has three sections: a list of the books included, a subject index to the stories, and a title listing of the stories.

The subject headings are based upon those used by Rue in her second edition of *Subject Index to Books for Intermediate Grades,* published in 1950. There are some additions to her subject list designed to add to its ready use. Cross references are plentiful.

The collections included are well selected and include those books one is likely to find in a well-stocked children's library. Some are out of print but were deliberately included because many libraries own copies of them. Grade levels for the collections are given.

In comparing this index with the ones by Rue and with the *Short Story Index* published by H. W. Wilson Company in 1953, some duplication of books indexed is found; about one-third of the books indexed in this volume being included in another index. This duplication may not be unwise since many teachers and librarians will not have access to all of the indexes mentioned. One note of warning to those using the list of books indexed as a finding list: they are arranged alphabetically by symbol, not strictly alphabetically by author and title. This detracts in no way from the usefulness of the book as an index, which is of course its intended purpose.

Everyone working in the field of literature for children will find this volume a most welcome addition to library collections.

Ruth E. Seeger

Runes, Dagobert D. *On the Nature of Man: an Essay in Primitive Philosophy.* New York: Philosophical Library, Inc., 1956. v + 105 pp.

This book provides a shock treatment for the complacent mind. The reader's arrogance and serenity will not receive one devastating blow, but hundreds of tiny, irritating rebuffs will dissipate any false contentment he may be cherishing. With a tedious array of smartly put aphorisms, euphemisms, and truisms the author follows what, to him, is a disappointing genesis of modern man. Law, God, reason, and other great ideals of the human intellect are depicted more often as the ingenious tools of crafty opportunists than as the discoveries of sacrificial, creative, philanthropic sages. The volume evinces little appreciation of the rôle of cultures, races,

governments, or technologies in the history of civilization. It oversimplifies the great environmental and social forces in mankind's march to freedom and dignity, and glibly contemns the significance and profundity of recent philosophical and scientific wisdom. As a work *On the Nature of Man*, it glaringly displays the follies and foibles of the human species, offering, at the same time, some moralistic, homiletical admonitions to those who live lives of honesty, kindness, and truth. For the reader to whom proverbs, platitudes, and pithy sayings are helpful and important, the thirty-three pages of "Evening Thoughts" will be stimulating bits for reflection.

D. Luther Evans

Hatchet, Ethel L., and Hughes, Donald H. *Teaching Language Arts in Elementary Schools: a Functional-Creative Approach.* New York: Ronald Press Company, 1956. xi + 426 pp.

These collaborators state that "the success of the functional-creative program of language learning is dependent upon the degree of relatedness it bears to the entire school program, which in turn must be based on the life needs and experiences of individual children" (page 328). The supportive background theory for such a program is well presented in the first four chapters. It is also well interpreted for total school and community programs in the final three chapters. In the nine intervening chapters the language-arts fields are covered with no slight to practical teaching methods, but with careful emphasis on developmental processes which recognize that children's natural experiences provide purposive settings in which learnings can be vital and lasting through the "functional-creative approach."

Those who share the reviewer's concern for helping children develop their listening powers will find Chapter VI a generator of ideas.

> Never before, in school or out, has there been such a rich opportunity for adventures in listening. But if the adventurers are to develop constructive and wholesome ideas for use in daily living, with power to think and to form good emotional attitudes, then more and better listening skills must be developed by the school (pages 128-29).

Along with providing important supplementary professional reading, recommended spelling lists, and reading lists for children, the authors have turned out a new kind of textbook with convincing information given for helping children through the various stages of

> . . . acquiring the *power to listen and to express* himself in words. . . . [achieving] purposeful *sharing or communicating* of ideas. . . . [to] *the potentially creative or open-ended stage.* . . . the stage of refining both the power to create and the ability to communicate ideas (pages 7-8).

Mary Jane Loomis